

(19) 日本国特許庁 (J P)

(12) 公表特許公報 (A)

(11) 特許出願公表番号  
特表2000-507377  
(P2000-507377A)

(43) 公表日 平成12年6月13日 (2000.6.13)

(51) Int.Cl. <sup>7</sup>	識別記号	F I	テームコード* (参考)
G 0 6 T 15/70		G 0 6 F 15/62	3 4 0 K
G 1 0 L 13/00		G 1 0 L 3/00	S

審査請求 未請求 予備審査請求 有 (全 43 頁)

(21) 出願番号 特願平9-534137  
(86) (22) 出願日 平成9年3月24日 (1997.3.24)  
(85) 翻訳文提出日 平成10年8月17日 (1998.8.17)  
(86) 国際出願番号 PCT/GB97/00818  
(87) 国際公開番号 WO97/36288  
(87) 国際公開日 平成9年10月2日 (1997.10.2)  
(31) 優先権主張番号 96302060.7  
(32) 優先日 平成8年3月26日 (1996.3.26)  
(33) 優先権主張国 ヨーロッパ特許庁 (EP)  
(81) 指定国 EP (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), AU, CA, CN, JP, KR, MX, NO, NZ, SG, US

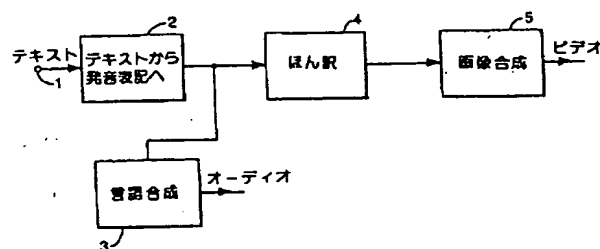
(71) 出願人 プリティッシュ・テレコミュニケーションズ・パブリック・リミテッド・カンパニー  
イギリス国、イーシー1エー・7エージェイ、ロンドン、ニューゲート・ストリート 81  
(72) 発明者 ブリーン、アンドリュー・ポール  
イギリス国、アイビー4・2ユーティアー、サフォーク、イプスウィッチ、ウエスターフィールド・ロード 50  
(72) 発明者 ボウアーズ、エマ・ジェーン  
イギリス国、ビーイー8・4エー2、ピーターバラ、アウンドル、パイン・クローズ 1  
(74) 代理人 弁理士 鈴江 武彦 (外4名)

(54) 【発明の名称】 画像合成

(57) 【要約】

(例えば合成言語を伴った) 顔の動画を合成することが入力フォニームストリングを一連の口の形すなわちビゼーム (Visemes) に変換して実行される。とくにある形が各母音に対してと子音を含んだ各遷移について生成される。

Fig.1.



**【特許請求の範囲】**

1. 話された言葉と整合がとれていて眼に見えるアーティキュレーションを備えた顔の動画を表わす信号を生成する方法であって、その構成は：

発話の継続している部分に対応する発音表記のシーケンスを受領することと；

第1の種類の各発音表記に対して口の形を識別することと；

第1の種類の発音表記から第2の種類の発音表記への各遷移と、第2の種類の発音表記から第1の種類の発音表記への各遷移と、第2の種類の発音表記から第2の種類の発音表記への各遷移とに対して口の形を識別することと；

該識別された形を含んでいる画像フレームのシーケンスを生成することを含む方法。

2. 話された言葉と整合がとれていて眼に見えるアーティキュレーションを備えた顔の動画を表わす信号を生成する方法であって、その構成は：

発話の継続するフォニームに対応する発音表記のシーケンスを受領することと

各母音フォニームに対する口の形を識別することと；

母音フォニームから子音フォニームへの各遷移と、子音フォニームから母音フォニームへの各遷移と、及び子音フォニームから子音フォニームへの各遷移とに対する口の形を識別することと；

識別された形を含んでいる画像フレームのシーケンスを生成することを含む方法。

3. 子音と母音とのフォニーム間の各遷移に対する口の形の識別が母音フォニームと子音フォニームとの関数として実行される請求項2記載の方法。

4. 2つの子音フォニーム間の各遷移に対する口の形の識別が2つの子音フォニームの第1のものと、その前後の直近の母音フォニーム関数として実行される請求項2又は3記載の方法。

5. 2つの子音フォニーム間の各遷移に対する口の形の識別が2つの子音フォニームと第1のものと、それに続く直近のもしくはそれがないときには先行する母音フォニーム関数として実行される請求項2又は3記載の方法。

6. 前記識別が請求項3, 4, 5のいずれか1項で特定されたフォニームだけ

の関数として実行されるその請求項記載の方法。

7. 前記識別が同じ単語内の少くとも1つの他のフォニームの関数としても実行される請求項3, 4, 5のいずれか1項記載の方法。

8. 各識別された口の形に対してその形を特定する指令を生成し、かつ前と後との指令によって特定された形の中間の形を各々が特定する中間指令を生成することを含む前記請求項1ないし8のいずれか1項記載の方法。

9. 話された言葉と整合がとれていて眼に見えるアーティキュレーションをもつ顔の動画を表わす信号を生成するための装置であって、その構成は：

動作時に発話の継続する部分に対応する発音表記のシーケンスを受領し、それに応答して第1の種類の各発音表記に対してと、口の形を識別し、また第1の種類の発音表記から第2の種類の発音表記への各遷移に対してと、第2の種類の発音表記から第1の種類の発音表記への各遷移に対してと、第2の種類の発音表記から第2の種類の発音表記への各遷移に対してとの口の形を識別するようにされた手段と；

識別された形を含む画像フレームのシーケンスを生成するための手段とを含んでいる装置。

10. 添付図面を参照して実質的に記述された、話された言葉と整合する眼で見る顔の動画を表わす信号を生成する方法。

11. 添付図面を参照して実質的に記述された、話された言葉と整合する眼で見るアーティキュレーションを備えた顔の動画を表わす信号を生成するための装置。

## 【発明の詳細な説明】

画 像 合 成

この発明は、例えば合成言語を伴う動く画像の合成に関する。

この発明によると、話された言葉と整合がとれていて、眼に見えるアーティキュレーション (articulation) を備えた顔の動画を表わす信号を生成する方法が提供されており、その構成は：

発話の継続している部分に対応する音声学上の表示（発音表記）のシーケンスを受領することと；

第1の種類の各発音表記に対して口の形を識別することと；

第1の種類の発音表記から第2の種類の発音表記への各遷移と、第2の種類の発音表記から第1の種類の発音表記への各遷移と、第2の種類の発音表記から第2の種類の発音表記への各遷移とに対して口の形を識別することと；

識別された形を含んでいる画像フレームのシーケンスを生成することとを含んでいる。

第1と第2の類型はそれぞれ母音と子音であってもよい。したがって、この発明の好ましい実施態様は話されたことばと整合がとれていて眼に見えるアーティキュレーションを備えた顔の動画を表わす信号を生成する方法であって、その構成は：

発話の継続するフォニームに対応する初音表記のシーケンスを受領することと；

各母音フォニームに対する口の形を識別することと；

母音フォニームから子音フォニームへの各遷移と、子音フォニームから母音フォニームへの各遷移と、子音フォニームから子音フォニームへの各遷移とに対して口の形を識別することと；

識別された形を含む画像フレームのシーケンスを生成することとを含んでいる。

。

子音と母音フォニーム間の各遷移に対して口の形を識別することは母音フォニームと子音フォニームとの関数として実行され、また2つの子音間の各遷移に対して口の形を識別することは2つの子音フォニームの第1のものと、それに続く

か先方する直近の母音フォニームの関数として実行されてよい。代って、2つの子音間の各遷移に対する口の形の識別は2つの子音フォニームの第1のものとそ

れを直近で続く母音フォニームか、それが無いときには先行する母音フォニームの関数として実行されてよい。

好ましいのは、各遷移に対する識別はこういった遷移と関係して上で特定したフォニームだけの関数として実行されることである。代って、識別は同じ単語内の少なくとも1つの他のフォニームの関数として実行することもできる。

好ましいやり方では、各識別された口の形に対して、その形を特定する指令と、中間の指令でその各々が前と後との指令によって特定される形の間の中間の形を特定する指令を生成してもよいようにする。

この発明の別な特徴では、話された言葉と整合する眼に見えるアーティキュレーションをもつ顔の動画を表わす信号を生成するための装置が用意されており、その構成は：

動作時に発話の継続する部分に対応する発音表記のシーケンスを受領し、それに応答して

第1の種類の各発音表記に対して口の形を識別し、また

第1の種類の発音表記から第2の種類の発音表記への各遷移と、第2の種類の発音表記から第1の種類の発音表記への各遷移と、及び第2の種類の発音表記から第2の種類の発音表記への各遷移とに対して口の形を識別するようにされた手段と；

該識別された形を含む画像フレームのシーケンスを生成するための手段とを含んでいる。

この発明の実施態様を、例を用いて、添付の図面を参照して記述して行く。

図1は、実施態様の要素を示す機能ブロック図である。

図2は、ヒトの頭部の画像を合成する際に使われる針金枠‘ワイヤフレーム’の平面図、正面図、及び側面図を示す。

図3は、ヒトの頭部の画像の口の部分を合成する際に使われる‘ワイヤフレーム’の同様な図である。

図4は、‘affluence’（アフルーエンス）と言うときのヒトの頭を表わすために画像のシーケンスを合成するとき最大の母音の口の形が生ずる場所を示す。図5は、同じ単語‘affluence’で最大の母音から子音（及びその逆）の遷移の

口の形が生ずる場所を示す。

図6は、同じ単語‘affluence’のアーティキュレーションで残っている口の形を示す。

図7は、同じ単語‘affluence’のアーティキュレーションで口の形の間の遷移を示す。

図8は、画像合成ユニットのための指令信号に発音信号をほん訳するためのユニットの部品を模式的に示す構成図である。

図9は、実施態様の装置の動作を示す流れ図である。

図10は、二重母音と破擦音を成分要素フォニームに変換するプロセスを示す流れ図である。

図11Aないし11Dは入力フォニームファイルに基いて中間出力ファイルを作るためのプロセスを示す。

図12は、中間出力ファイルに基づいて最大の口の形の本質とタイミングとを特定するファイルを作るためのプロセスを示す。

図13Aと13Bとは最大の口の形と中間の口の形との両方を特定するファイルを作るためのプロセスを示す。

図1の装置は、話されることになる単語を、テキストの形で受領し、対応する言語をオーディオ信号の形で生成し、かつ顔の動画の表示のための対応するビデオ信号を生成し、同じ言語と対応している口のアーティキュレーションをもった顔（例えばヒトの顔であったり漫画であったりする）とするような機能を有している。ここでの記述にあたっては、口のアーティキュレーションとしばしば言うがこのアーティキュレーション（特定の発音をする際に含まれる全体としての発声器官の調節や動きを指す：articulation）は唇、口の内部（ときとして歯と舌とを含めて）、あご、及びその周辺の動きも含んでいると理解されたい。例えば全体としての頭部の動きや回転、眉の動きといった他の動きもまた、結果として

得られる画像を一層リアルなものとするために含まれていてよい。

記憶されているテキストファイルもしくは他の所望のファイルからのテキストが、入力1でいずれかの便利な規格表示（例えばASCIIコード）に従った文字コードの形態で受領される。これは通常構成の言語合成器によって受領される

が、ここでは二つの別な部品、すなわち、テキストから発音表記への変換器2は通常の綴り字を発音表記、例えばフォニームのリストと各々の継続時間に変えるものとして、また言語合成器だけのもの3はこのリストをオーディオ周波数波形に変えるものとして示されている。いずれのフォニーム組も使用できるが、ここでの記述の目的ではイギリスのRP-SAMPA組の使用をして、場合には次の表1に記したイギリス英語（British English）の38の個別フォニームを識別すると仮定している。

表1

イギリスの RP-SAMPA	単語例
子音	
/b/	be <u>ar</u>
/D/	th <u>i</u> s
/d/	de <u>a</u> r
/f/	fe <u>a</u> r
/g/	ge <u>a</u> r
/h/	he <u>a</u> r
/j/	ye <u>a</u> r
/k/	ki <u>n</u> g
/l/	le <u>a</u> d
/m/	me <u>n</u>
/N/	wi <u>n</u> g
/n/	ne <u>a</u> r
/p/	ge <u>a</u> r
/r/	re <u>a</u> r
/S/	she <u>e</u> r
/s/	si <u>n</u> g
/T/	thi <u>n</u> g
/t/	te <u>a</u> r
/v/	ve <u>r</u> y
/w/	w <u>e</u> ar
/Z/	tre <u>a</u> sure
/z/	zo <u>o</u>
破擦音	
/dZ/	je <u>e</u> r
/tS/	che <u>e</u> r

イギリスの RP-SAMPA	単語例
短母音	
/@/	ag <u>o</u>
/I/	ba <u>t</u>
/E/	be <u>a</u> t
/I/	bi <u>t</u>
/Q/	co <u>d</u>
/U/	go <u>o</u> d
/V/	bu <u>d</u>
長母音	
/3/	bi <u>r</u> d
/A/	ba <u>r</u> d
/i/	be <u>a</u> d
/O/	bo <u>r</u> e
/u/	bo <u>o</u> t
二重母音	
/@U/	ze <u>r</u> o
/aI/	pie
/aU/	co <u>w</u>
/E@/	ha <u>i</u> r
/eI/	pa <u>y</u>
/I@/	pe <u>e</u> r
/OI/	bo <u>y</u>
/U@/	co <u>n</u> to <u>u</u> r
ほか	
/#:/	無音 (静寂)
/#/	単語境界

音声合成器は通常のものであり、ここではもっと記述することをしない。

フォニームリストはほん訳ユニット4で受領され、それについては以下でもっと詳細に記述する。このユニット4は、フォニームリストから、一連の指令信号を生成し、顔に必要とされる口のアーティキュレーションを特定して、フォニー



ムリストに対応したやり方でそれが動くようにし、それによって合成器3により生成された言語信号と対応する動きとなるようにする。

これらの指令信号は画像合成ユニット5によって受領される。このユニットは単一ビデオフレームもしくは所望の顔の静止画のビットマップ画像をその中に記憶していて、この顔を示す連続したビデオ信号を、動きを伴って生成するように働く。はっきりしていることはこのビデオ信号が希望するどの規格のものでもよく、ここでは毎秒25フレームをもつシステム1 (System 1) 信号を仮定している。動きは三次元ワイヤフレームモデルの助けをかりて生成される。典型的なこの種のモデルは図2に示されており、口の部分については図3で拡大して示されている。三次元空間内には多数の点(頂点)があり、またこれら頂点を結ぶ直線がファセット(facet)と呼ばれる三角形の領域を定義している。実際の装置では、記憶されたデータの組としてモデルが存在し、言い換えると、各頂点に対して頂点番号とその $x$ 、 $y$ 、 $z$ 座標とがあり、また各ファセットに対してファセット番号とそのファセットの隅を形成する3つの頂点の番号とがデータとして記憶されている。初期化段階では、ユニット5はこの基準モデルの各ファセットとビットマップ画像の対応する領域との間の写像を判断する。変更されたモデルを繰返し定義することによって動きが作り出され、変更されたモデルでは1又は複数の頂点が基準モデル内で占有していた位置から別な位置に行っていると想定する。ユニット5はそこで新しい二次元のビットマップ画像を生成する必要がある。これをするのは、変更したモデルのいずれかのファセットすなわち基準モデルに対して動いた1又は複数の頂点を識別することにより、このようなファセットの各々に対して、補間プロセスを採用し、このプロセスではもとのビットマップの三角形領域がマッピングによって移動もしくはひずむかあるいは両方をして新しいビットマップ画像内で三角形領域を占有し、このマッピングによって変更したモデルのファセットに対応がとれる。このような新しいビットマップが出力信号の各フ

レーム(すなわち毎40ミリ秒)に対して生成される。画像合成ユニット5の動作と実施についてももっと詳しいことについては、W. J. Welsh, S. Searby and J.

B. Waite "Model Based Image Coding", Brit. Telecom. Tech. J. vol8, No. 3, July 1990を参照とする。

画像合成ユニット5を駆動するために必要とされる指令は、原理上、40ミリ秒毎に、基準モデルとは違う位置の各頂点の番号を、その新しい座標と一緒にユニット5に送れることになっている。しかしながら関心のある動作速度では、ユニット5は動作ユニットの記憶された組を含んでおり、その各々はデータエントリが次の構成のものとなっている：

—動作ユニット番号（例えば0ないし255）（1バイト）

—この動作ユニットで影響される頂点の番号

—このような各頂点に対して：

その頂点の番号（2バイト）

基準モデル内のその位置からのx座標変位（2バイト）

基準モデル内のその位置からのy座標変位（2バイト）

基準モデル内のその位置からのz座標変位（2バイト）

（もしその方がよければ、無論前のフレームに対するx, y, zのシフトが使える）。

そこで各指令は単に動作ユニット番号に、この動作ユニットによって特定される動きの量を変えるためのスケーリング因子（例えば0ないし255）を追従させたもので構成されてもよい；あるいはもし望むのであればいくつかを含んでもよい（試作段階では最大5が可能であった）。ユニット5は指令を受領すると、動作ユニットを一覧対照し、記憶している座標シフト（適切にスケールが作られているものとする）を特定の頂点に対して使用する。もし指令が2つの動作ユニットを含み、その両方が特定の頂点の変位を特定するのであれば、この変位は2つの変位の単なるベクトル和である。

ほん訳ユニット4の動作についてここで見ることにし、visemeという概念を導入することが便利である。話された単語がフォニームと呼ばれる要素単位（elemental unit）で構成されると見ることができるのと全く同じに視覚言語（visual

speech）もビゼーム（viseme(s)）：phonemeを音から視覚visionに置きかえた用語

）で構成されていると見ることができる。すなわち、四角言語の最小単位であり、目視可能なアーティキュラトリイ（調音）ユニットの最小認識単位である。基本的には、ビゼームは口の形であり、ほん訳ユニットの課題は、どんなビゼームが要求されているかと、それらがいつ発生するかという時刻とを判断して、40ミリ秒間隔で指令を発生し、さらに、必要とされる間隔で必要とされるビゼームを生成し、間に入るフレームのために適当な中間形状を生成することである。

ほん訳ユニットの動作に重要なのは、フォニームとビゼームとの間に1対1の対応が存在しないという考え方である。まず、なにがしかのフォニームは視覚的に類似し、ときには区別が不能である。たとえば、子音／p／と／b／とは視覚的には同一であり、発生が程度が違っただけで、発生器官のアーティキュレーションは同じである。したがって、フォニームは群形成をすることができ、同じ群のフォニームではビゼーム生成に関する限りは同一と考えられる。いろいろな群形成が可能で、典型的な群形成を以下の表2に示す。

表2

フォニーム	群
p, b, m	子音群 1
f, v	子音群 2
D, T	子音群 3
s, z	子音群 4
S, Z	子音群 5
k, g, N	子音群 6
t, d, l, n, r	子音群 7
w, U, u, O	"b o t h (両方)"群
Q, V, A	母音群 1
3, i, j	母音群 2
@, E, I, {	母音群 3

（注：二重母音がないが、処理前に構成要素母音に分割されるためである）

第2に、母音の音と口の形との間の関連性を定義することは可能であるが、子音についてはそうはならず、子音では口の形が近くのフォニームに依存して変り

、とくに母音のフォニームの近くで変る。この実施態様では口の形は母音と、子音とフォニームとの組合せとの両方に関連性がある。子音を含む遷移にはかなりの数が存在する。しかし、第1の簡略化が可能であり、子音から子音への遷移が後続の母音（もしくはポーズの前の単語の終りでは先行する母音）による大きな影響を受けていることを観察することによってされ、2つのうちの第2の子音は若干の効果はあるが、全くぼんやりしたものでなく、無視できるものである。この実施態様は各子音から子音の遷移に小音－母音もしくは母音－子音の組合せを関連づけることによりこの利点を採用している。こうして、システムによって処理される必要がある口の形の数を少く保っている。

例を用いてこの実施態様の動作を示すために、もしテキストから発音表記へ変換するユニット2が単語‘affluence’を表わす信号を受けたとすると、このユニットはフォニームリスト/#:/ / [ / / f / / l / / u / / @ / / n / / s / / #:/をほん訳ユニット4に出力することになる。このフォニームリストを受領するとほん訳ユニット4はフォニームリストを処理するように動作して一連の指令信号を出力する。出力指令信号は図4ないし7に示されていて、その各々はまた入力フォニームリストの内容を、すなわちフォニーム自体とその継続期間をサンプルで（この例ではサンプルレートは8kHzである）示している。

まず、この出力には3つの指令信号でこの単語内の母音に対応するものが含まれている。これらが図4に示されていて、下側の図は母音/ [ / / u / 及び / @ / が識別され、各々はその母音に割当てられたビゼームが判断されたことを示すバー（棒）でしるしをつけられていて、母音の midpoint で生じるようにとられている。

出力はさらに母音－子音と子音－母音遷移と関連する口の形を特定する指令信号を含んでいる。これが図5に示されていて、ここではバーは母音－子音もしくは子音－母音の境界における口の形を示している。これは子音－子音遷移を残している。前に述べたように、この遷移1は主として第1の子音とそれに続く子音

とによって特徴づけられるものとして見られている。したがって、/f/から/l/への遷移が図6で表わされていて、子音－母音組合せ、/f/から/u/、に対する口の形となっている。/n/から/s/への遷移は次に続く母音がなく、それ故に使

われる口の形は/@/から/s/の母音-子音組合せに、すなわち先行する母音の使用、に対応したものである。先行と後続の無音（静寂期間/#:/は無論、閉じた口をもつ（すなわち、基準ワイヤフレームモデルの）顔で表わされる。

図6でバーでしるしをした時刻の瞬間には（あるいはこれらの瞬間の一番近い40ミリ秒期間では、ほん訳ユニット4は画像合成ユニット5に対して問題となっている口の形に適したスケーリング因子と動作ユニットとを特定する指令を送る。これらの瞬間の間の40ミリ秒では、2つの口の形の中間の口の形を特定する指令を送る必要がある。例えば、{fとマークを付けた瞬間とfuとマークを付けた瞬間との間では2つの動作ユニットを特定する指令を送り、それがそれぞれ母音-子音組合せ、/[/から/f/へと、子音-母音組合せ、/f/から/u/へ、とに対応しており、縮小したスケーリング因子には無関係であり、それよって2つの形の間で滑らかな遷移が得られるようにする。したがって、2つの瞬間の間の途中の点x%では、組合せ、/[/から/f/、のための動作ユニットにはスケール因子としてそのスケール因子の $(1-x/100)$ 倍が[f点で送られ、それと一緒に組合せ/f/から/u/のための動作ユニットにはスケール因子としてそのスケール因子の $x/100$ 倍がfu点で送られる。図7はこのプロセスを図式的に示している。中間コマンド信号を作るという目的には無音フォニームと関連の口の形は無音フォニームの中心が到達する前の後続の口の形によっては影響されていないことが分かる。

上記表2の11の群については7つの子音群があり、また3つの母音群と1つのいわゆる“両方の”群とがあった。この“両方の”群には母音フォニームと子音フォニームとが含まれている。したがって、無音を含む遷移を無視すると、必要とされる母音と、母音-子音及び子音-母音組合せのすべては母音群と、母音群-子音群及び子音群-母音群の組合せで次の表3に示すものによって表わすことができる。

表3

母音	4
子音群から母音群への組合せ	21

母音群から子音群への組合せ	21
両方の群から他の群への組合せ	10
他の群から両方の群への組合せ	10
両方の群から両方の群への組合せ	2
合計	<u>68</u>

これら68の母音群と群の組合せの若干のものは同じ口の形に対応している；さらに若干の口の形は他のものと似ており、主としてプロポーションに違いがある。換言すれば同じ動作ユニットにより作ることができるが、違ったスケーリング因子を備えている。（後述するところであるが、）動作ユニットの判断の間に、68の母音群と群の組合せとは1:1の動作ユニットと適当なスケーリング因子で表わされることが見付かった。表4はこれらを記述したもので、動作ユニットについての記述と、スケーリング因子とともに増大する特徴についての記述と、この動作ユニットによって表わすことができる母音群と群の組合せとのリストも添えてある。与えられた母音群と群の組合せに対応するそれぞれの口の形を作るのに使われることになるスケーリング因子も示されている。

当業者であれば、多数の動作ユニットを定義してもよいことが分ると思うが、その場合、母音群と群の組合せとは動作ユニットの間でもっと精細に分けられることになる。

表4

動作 エッセ 番号	記述	母音群又は群組合せ	スケール
1	丸い口、 突出した唇、 歯も一緒。 口の形は スケールで 一層丸くなる。	母音群1から子音群5へ	125
		母音群2から子音群5へ	130
		母音群3から子音群5へ	125
		“両方の”群から子音群5へ	120
		子音群5から母音群1へ	120
		子音群5から母音群2へ	120
		子音群5から母音群3へ	125
		子音群5から“両方の”群へ	120
2	歯は使わず、非常に 丸い唇の外形線、唇 の間の隙間は直線だ が小さい。 口の形はスケールで 一層丸くなる。	“両方の”群から母音群2へ	150
		“両方の”群から母音群3へ	150
		“両方の”群から“両方の”群へ	150
		“両方の”群から“両方の”群へ	130
		“両方の”群から子音群7へ	120
		子音群7から“両方の”群へ	120
3	長い口の形、 上歯のみ、 下唇はしまい込む、 歯がスケールで突出 する。	母音群1から子音群2へ	100
		母音群2から子音群2へ	110
		母音群3から子音群2へ	115
		“両方の”群から子音群2へ	100
		子音群2から母音群1へ	100
		子音群2から母音群2へ	100
		子音群2から母音群3へ	115

		子音群2から“両方の”群へ	100
4	口の形は長くかつ丸め、歯は使わず、唇間の隙間は丸い。唇間の隙間はスケールで一層大きくなる。	母音群1から“両方の”群へ	240
		母音群1から子音群1へ	130
		母音群2から“両方の”群へ	240
		母音群3から子音群1へ	130
		“両方の”群	130
		“両方の”群から母音群1へ	240
		“両方の”群から子音群3へ	130
		“両方の”群から子音群7へ	130
5	動作ユニット4と同じ、ただし上唇はもっと曲っている。	母音群1	130
		子音群1から母音群1へ	95
		子音群1から“両方の”群へ	80
6	長い口の形、上と下の歯が見えるがその間の隙間は見えぬ。隙間はスケールで一層大きくなる。	母音群3から子音群6へ	110
		“両方の”群から子音群6へ	110
		子音群3から母音群2へ	130
		子音群6から母音群3へ	110
7	丸めた口の形、上と下の歯が見えるがその間の隙間は見えぬ。隙間はスケールで一層大きくなる。	母音群2から子音群6へ	110
		母音群3	140
		子音群6から母音群1へ	130
		子音群6から“両方の”群へ	110
		子音群7から母音群1へ	130
	長く僅かに丸めの口の形、上の歯は見え	母音群2	160
		母音群2から子音群6へ	160



9	る。上の歯はスケールで一層突出する。	子音群 4 から母音群 3 へ	170
		子音群 6 から母音群 2 へ	160
		子音群 7 から母音群 3 へ	170
		子音群 7 から母音群 3 へ	125
11	長い口の形、上の歯は見える。上の歯はスケールで一層突出する。	母音群 3 から子音群 4 へ	130
		母音群 3 から子音群 7 へ	120
		“両方の” 群から子音群 4 へ	105
		子音群 4 から “両方の” 群へ	105
12	動作ユニット 11 と同じ、ただし上唇はそんなに丸めていない。	母音群 1 から子音群 4 へ	100
		母音群 1 から子音群 7 へ	100
		母音群 2 から子音群 4 へ	120
		母音群 2 から子音群 7 へ	120
		子音群 4 から母音群 1 へ	130
		子音群 4 から母音群 2 へ	110
13	長い口の形を上歯と舌で。 歯はスケールで一層突出する。		120
		母音群 1 から子音群 3 へ	105
		母音群 2 から子音群 3 へ	110
		母音群 3 から子音群 3 へ	115
		子音群 3 から母音群 1 へ	105
		子音群 3 から母音群 2 へ	105
		子音群 3 から母音群 3 へ	130
		子音群 3 から “両方の” 群へ	105

ほん訳ユニット 4 は適当にプログラムされた処理ユニットという手段で実現され、これが図 8 ではプロセッサ 10、プログラムメモリ 11、及び多数のメモリでルックアップ表を含むもので構成されている。とくにここには二重母音表 12、フォニーム群表 13、及び動作ユニット表 14 を含んでいる。明りょうにする

ためにこれらが別に示してはあるが、無論、単一のメモリが実際にはプログラムとルックアップ表とを含むことができる。メモリ11内に記憶されているプログラムの動作を図9ないし13に示す流れ図を参照してこれから詳細に説明して行く。

図9の流れ図は全体としての装置の動作を簡単に示しており、図10ないし13によって表わされるアルゴリズムがそこで発生するようなコンテキストを設定している。このアルゴリズムはプログラムメモリ11内に記憶されていて、動作ユニットファイル（動作ユニットとスケーリング因子とを含む）を生成するように実行でき、このファイルは画像合成ユニット5に送られることになる指令信号に対する基礎を形成している。したがって、段階100の初期化に続いて、言語合成器のテキストから発音表記へのユニット2によって受領されたテキストメッセージは段階104でフォニームファイルを作る。このファイルの受領がほん訳ユニット4で認識されると（段階106、Text-To-Speech）、ほん訳が行なわれ（段階108）でフォニームリストが動作ユニットファイルとされる（段階110で作られる）。これが画像合成ユニット5へ（段階112で）送られる指令信号に対する基礎を形成し、同時にフォニールファイルが合成器3に送られる。もし望めば無音（静寂）の間に（段階114）もしくは言語の間に（段階116）、追加の動作ユニットがランダムな（もしくは他の）頭部の動きを作ることができる。

段階108の動作は図10に示した流れ図によって示されたプログラム段階を用いて二重母音と破擦音の拡張をすることで始まる。このプログラムはフォニームファイルの各要素を順に読み（段階120）、2つの文字によってそのフォニームが表されているかどうかを判断する（段階122）。もしそうであれば、このプログラムはプロセッサが要素をその構成要素文字に分けて、それらの文字によって表わされる2つのフォニームで要素を置換える。各々の継続期間は分けられた二重母音又は破擦音フォニームの継続期間の半分に設定される。フォニーム

出力のリスト内のフォニームの数を測定する可変の（noofphoneme：フォニームの数を意味する）は1だけ増分される（段階126）。そうでなければ、この要

素はフォニームリストに加えられる（段階128）。

二重母音表12の助けを得て/a I/, /a U/, 及び/e I/のような二重母音をフォニーム対/ [+I/, / [+U/, 及び/E+/I/にそれぞれ変換することが例示のプログラム段階で実行可能なことが分ると思う。同様に、このプログラムは破擦音/d Z/と/t S/とを2つのフォニームに分けるように実行できる。

次に要素毎に図10に示したプロセスで作られたフォニームリストの検査が続く（図11A-11D）。最初の無音フォニームの後の各要素に対して、フォニームの組合せもしくは母音と関係する時間間隔とが中間の出力フィル内に記録される。したがって、各エントリはフォニーム組合せもしくは母音で前の口の形の瞬間と現在の口の形の瞬間との間で作られることになるものを時間間隔と一緒に識別する（すなわち、この時間間隔は図6のバーの間の距離に対応している）。以下に別段の記述をするほかは、各エントリの後にはプログラムは判断段階180に戻って、フォニームリストの最終要素に到達しているかどうかを判断する。もしそうであれば、フォニームリストの検査は終る。もし到達していなければ、プログラムは現在の要素の分類段階130に戻る。

フォニームリストを検査するために、各要素に対してその要素が母音であるか、子音であるか、または無音であるかが判断される（図11A-段階130）。

現在の要素分類段階130で母音が見付かるとすると、図11Bに示した段階が実行される。まず、フォニームリストの中で前のフォニームが無音か、子音か、母音かを見付ける（段階140）。前のフォニームが無音フォニームであると、そのときは母音の口の形の前の時間間隔が母音継続期間の半分と無音継続期間の半分との和に設定される（段階141）。無音から母音への遷移がそこで計算された時間間隔と一緒に中間出力ファイルに入れられる（段階142）。もし前のフォニームが母音フォニームであると、そのときは母音の口の形の間の時間間隔は現在の母音の継続期間の半分と前の母音の継続期間の半分との和に設定される（段階143）。再び、母音自体（例えば/@/）と関連する時間間隔がそこで

中間出力ファイルに入れられる（段階144）。もし前のフォニームが子音フォニームであると、そのときは前のフォニームの前のフォニームが無音かどうか

判断される（段階145）。そのときは、前の口の形からの時間間隔が現在の母音の継続時間の半分に設定され（段階146）、その母音が計算された時間間隔と一緒に中間出力ファイルに入れられる（段階147）。もしそうでなければそのときは、前の口の形からの時間間隔は子音の継続時間に設定され（段階148）、子音から母音への組合せ（例えば/ l /から/ u /）と関連の時間間隔とが中間出力ファイルに入れられる（段階149）。この点で、プログラムは判断段階180に戻らずに、遷移ファイル内に別のエントリが行なわれるようにされ（段階146、147）、このエントリには現在の母音と母音自体（例えば/ u /）の継続時間の半分に等しい時間間隔を含んでいる。

図11Bの段階のもつ1つの効果は現在の母音に対応する口の形が母音フォニームの中央と一致することを確認にすることである。

現在のフォニーム分類段階（130）で無音が見付かるときは、図11Cの段階が実行される。まず、フォニームリスト内の前のフォニームが無音か、子音か母音かが見付けられる（段階150）。もし前のフォニームが無音であると、そのときは誤りが表示される（段階151）。もし無音の前に母音があれば、そのときは前の口の形からの時間間隔が母音の継続期間の半分と無音の継続期間の半分との和に設定され（段階152）、母音から無音への遷移が中間出力ファイル内に時間間隔と一緒に記録される（段階153）。もし前のフォニームが子音であれば、そのときは最後の口の形からの時間間隔が子音の継続期間と現在の無音の継続期間の半分とに設定される（段階154）。この場合、母音-子音組合せから母音への遷移（例えば/ @ s /から/ # : 1 /）と関連の時間間隔とが中間出力ファイルに入力される（段階155）。

もし段階130で子音が見付かると、図11Dに示した段階が実行される。まず前のフォニームが母音、無音、もしくは子音として分類される（段階160）。もし母音であれば、そのときは時間間隔が母音の継続時間の半分に設定され（段階161）、母音-子音の組合せ（例えば/ [ /から/ f /）は時間間隔と一緒に中間出力ファイル内に記録される（段階162）。前のフォニームが子音であ

れば、そのときはプログラムが母音フォニームについてフォニームリストの前方

探査を行なう（段階163）。もし子音－母音組合せ（前の子音と後の母音の組合せ）（例えば/t/から/u/へ）と関連の時間間隔（前の子音の継続期間に等しい）が中間出力ファイル内に入れられる（段階164, 165）。前方探査で母音が見付からなければ（段階163）、そのときはプログラムはプロセッサに母音の後方探査をさせる（段階166）。もしこの探査が成功すれば、そのときは母音－子音組合せ（初期の母音と現在の子音のもので、例えば/@/から/s/へ）が関連する時間間隔（前の子音の継続期間に等しい）と一緒に記録される（段階167, 168）。もし、前方探査も後方探査もともに母音を見付けないと、誤り表示が生ずる（段階169）。現在の子音に直近の先行フォニームが無音であるとして見付かると、そのときは母音に対する前方探査が行なわれる（段階170）；もし母音が見付かると、現在の子音と先行する無音の継続期間の半分との和に等しい時間間隔が無音から子音－母音組合せへの遷移と一緒に中間出力ファイル内に記録される（段階171, 172）。もし母音が何も単語内で見付からないときは誤りが表示される（段階173）。

図12では、母音とフォニームの組合せで中間出力ファイル内にあるものがlookupアップ表13にアクセスして母音群とフォニーム群の組合せに変換される。原理的にはこの内容は上述の表2内に設定されたようなものであり、したがって、各母音もしくはフォニームの組合せが群番号に変わる。しかし、各群を群番号で表わすのではなく、その群の1つの指定されたフォニームによって表わすのがもっと便利であることが見付かっている。例えば、フォニーム/p/, /b/, 及び/m/は全部が/p/に変えられる。これをするためには、プロセッサは図12に示したプログラムによって制御される。中間出力ファイル内の各要素に対して、要素の類型が次の1つであると判断がされる（段階190）：母音（段階192が実行される）；母音／子音組合せ（段階194が実行される）；母音／無音遷移（段階196が実行される）；あるいは組合せから無音への遷移（段階198が実行される）。段階（192, 194, 196, 198）は各構成要素母音もしくは子音を、群を表わすために選ばれた母音または子音に変換することができる。このプロセスは群／群組合せリストに戻るが、今では上述のように、最大6

8の異なる母音群と子音群の組合せを含んでいる。

図13Aと13Bでは、結果として得られる群リストが動作ユニットルックアップ表14を用いて動作ユニットファイルに変換され、群／群組合せリスト内の各要素を表わす動作ユニットを見付けるようにする。（ルックアップ表14の内容は上記表3のコラム3、1及び4内に設定されているものであるか、あるいはもし好ましい選択肢であれば、コラム3内の代表的なフォニームを伴うものとする。）この動作ユニットファイルはそこで40ミリ秒間隔で生成される指令信号のシーケンスを発生するために使用される。

もっと詳しく述べると、変換プロセスは群リストから第1の要素をフェッチすることで始まり（段階200）、その後動作ユニットルックアップ表はその要素に関連する動作ユニットとスケーリング因子とを決める（段階201）。次に第1の要素に関連する時間間隔内で完全な40ミリ秒期間の数を計算する（段階202）。初期動作ユニットのスケーリング因子が次に期間の数で除算されて増分値を与えるようにする（段階203）。このプロセスは次に命令のループ（段階204）に入り、各40ミリ秒期間に対する指令信号を作る。指令信号内のスケーリング因子は命令のループが実行される度毎に（ゼロから）計算された増分だけ増加される。

群リスト内の次の要素がそこでフェッチされ（図13Bの段階205）、対応する動作ユニットとスケーリング因子とが動作ユニットルックアップ表14を用いて見付けられる（段階206）。段階202のように、群リストの要素に関連する時間間隔内の40ミリ秒期間全体の数が見付けられる（段階207）。前のように、現在の要素に関連する動作ユニットのスケーリング因子が計算された期間の数で除算されて増分値が求められる（段階208）。群リスト内の前の要素のスケーリング因子が同じ数で除算されて減分値が求められる（段階209）。このプロセスは次に命令のループに入って、出力すべき指令信号を計算する。これには前の要素との関係で作られた動作ユニットと群リスト内の現在の要素に関連する動作ユニットとの重みづけした組合せを含んでいる。前の動作ユニットに与えられる重みづけは各40ミリ秒期間に対する減分値だけスケーリング因子を減分することにより減らされ、現在の動作ユニットに与えられる重みづけは各4

0ミリ秒期間に対する増分値だけ（ゼロから）スケーリング因子を増すことにより増加される。このようにして、指令信号出力は1つの口の形から次へと段階的な遷移して行く。

同じような動作（段階206ないし210）が群リスト内の後続の各要素に適用され、最終要素に到達するまで進む。

指令信号が動作ユニットファイルに基づいて生成されて、40ミリ秒間隔で画像合成ユニット5に送られて、頭部の画像を生成できるようにし、頭部にはテキストから言語への合成器の出力に対応しているアーティキュレーションを備えるようにする。

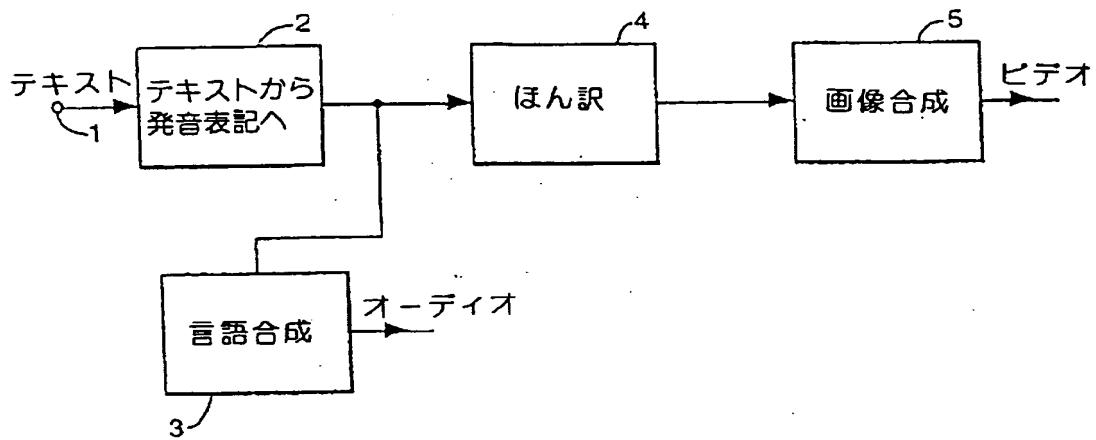
これまでの議論から、ビゼームあるいはある母音に対して選ばれた口の形は、その母音に先んじて選ばれたものであること、母音ー子音（もしくはその逆）の組合せに選ばれた口の形はその組合せに対して先行して割当てられたものであること、また子音ー子音遷移のために選ばれた口の形は同じ文脈（コンテキスト）内の子音の第1のものに先行して割当てられたものであること一言い換えると、この例では同じ次に来る母音（あるいはデフォルトであるか、先行している母音）をもつものであることが気付くと思う。もし望むのであれば、必要とされる動作ユニットの数を増すという負担を伴うけれども、口の形の選定はもっと文脈に依存するようにしてもよい。例えば、子音ー母音遷移に対する口の形を選ぶのに、子音とその次の母音とだけ依存するのではなく、前の母音にも依存するように（すなわち母音ー子音ー母音の組合せに依存するように）選択してもよいことになる。子音ー子音遷移に対する選択は第1の子音と後続及び先行の両母音（もしあれば）に依存するようにできるし、または実際に2つの子音と2つの母音に依存するようにできる。

これまでには、動作ユニットが画像合成ユニット5内でどのように生成されるかについて余り触れるところがかかった。試作品ではこれが達成されており、必要とされる68の母音群と母音群／子音群組合せの全部を含んだ人間の話した単語をビデオ記録し、かつフレームグラバ（獲得手段）を用いてこの記録の静止画フレームを表示して、これらのフレームで母音に対応するものと、子音／母音組合せに対応するものとを人手によって識別できるようにした。一度これらのフレ

ームが（ビットマップ形式で）識別されてしまうと、次はこれらのフレームを表わす基準ワイヤフレームモデルからの変位を判断する必要があった。これは一致プログラムを用いて行われ、このプログラムはワイヤフレームモデルを与えられたビットマップ画像に一致させるのに必要な変形を計算するものである。

【図1】

Fig.1.





【図2】

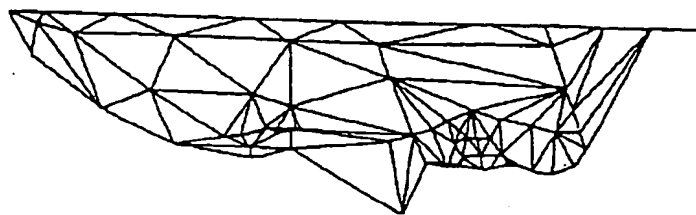
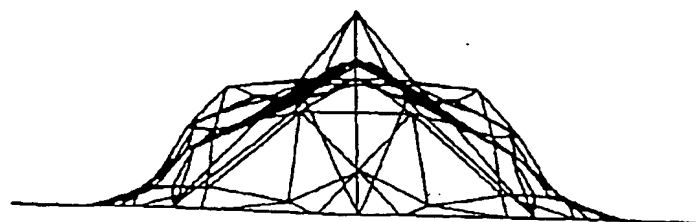
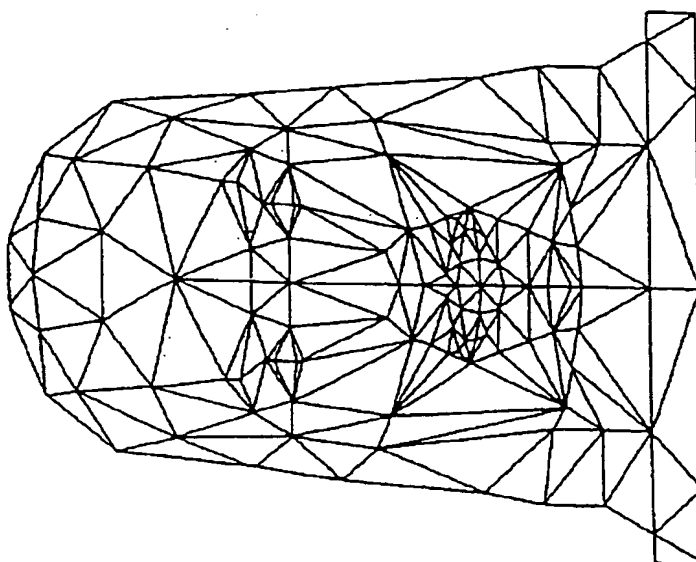
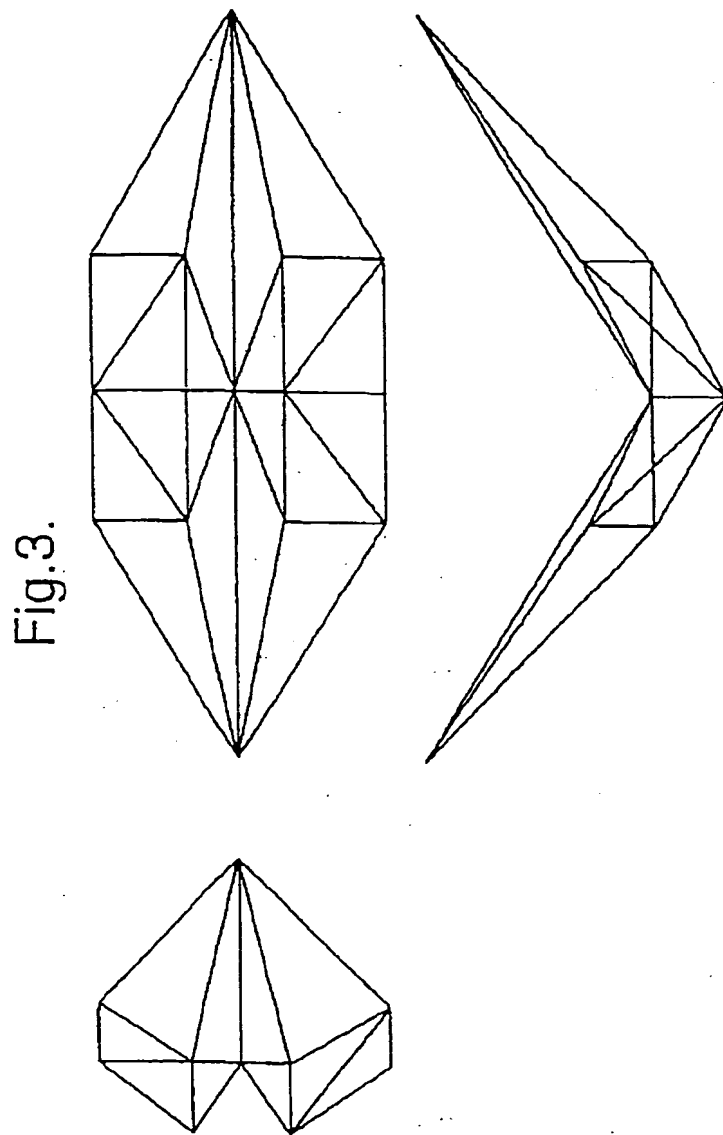


Fig.2.

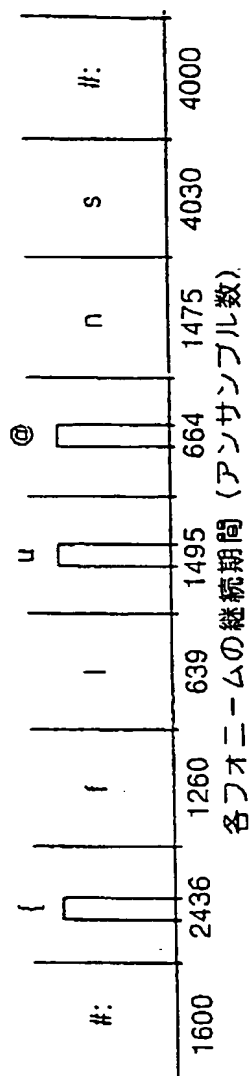
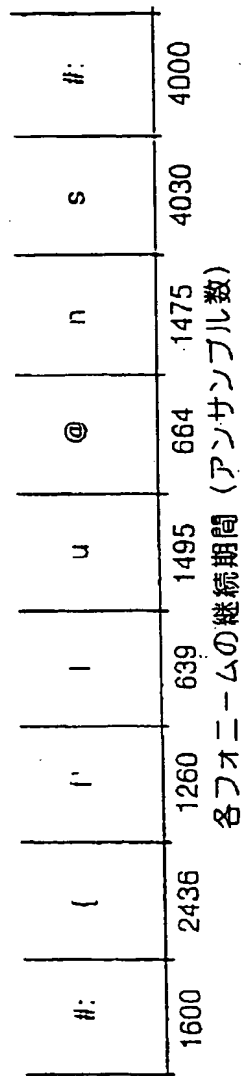


【図3】



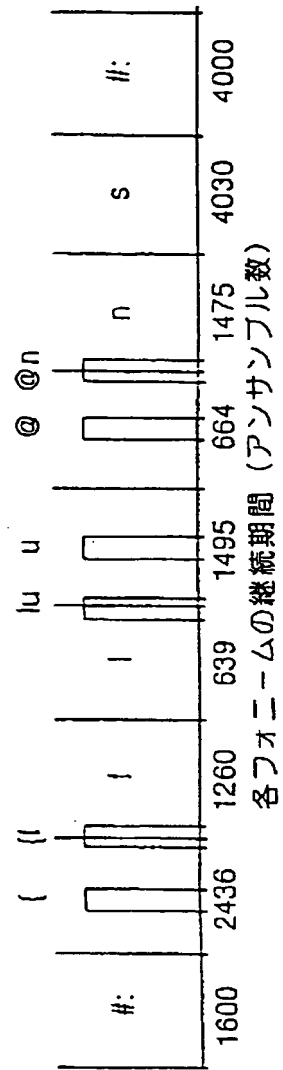
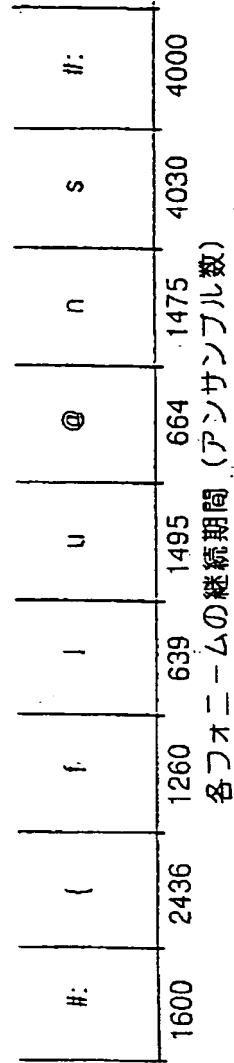
【図 4】

Fig.4.



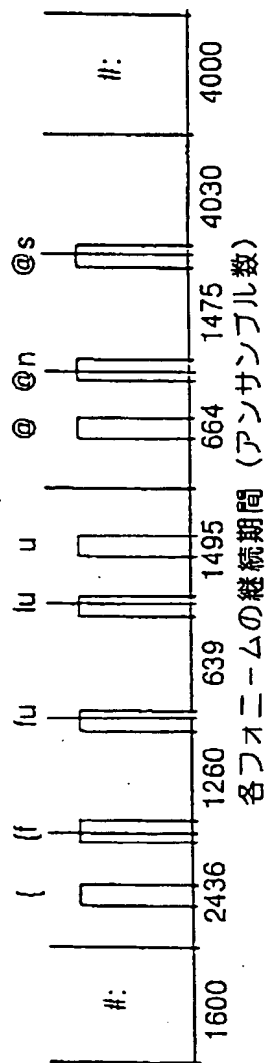
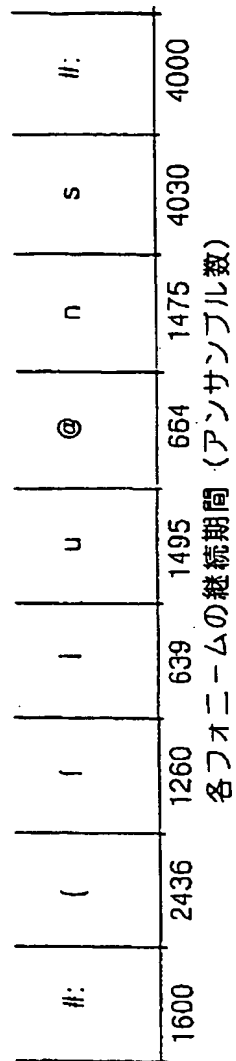
【図5】

Fig.5.



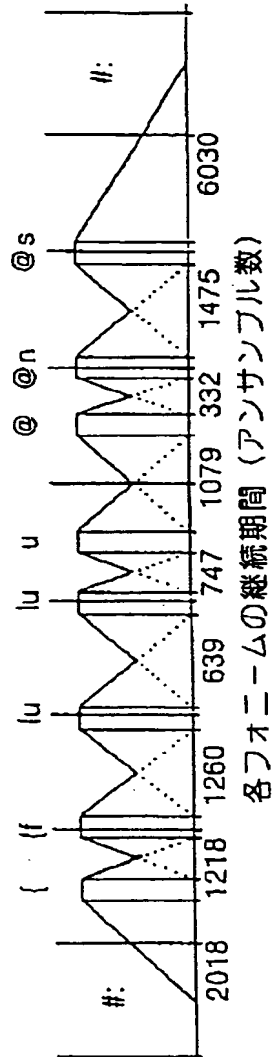
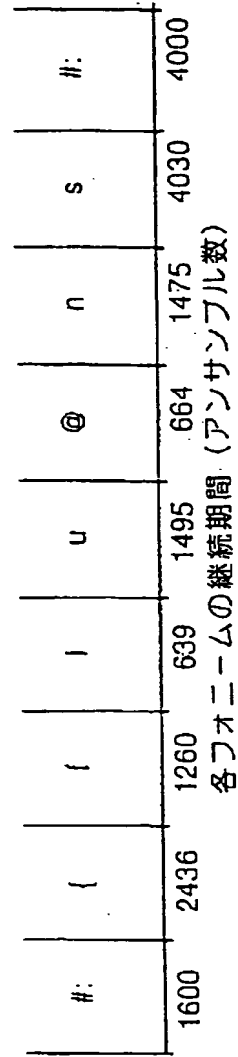
【図6】

Fig.6.

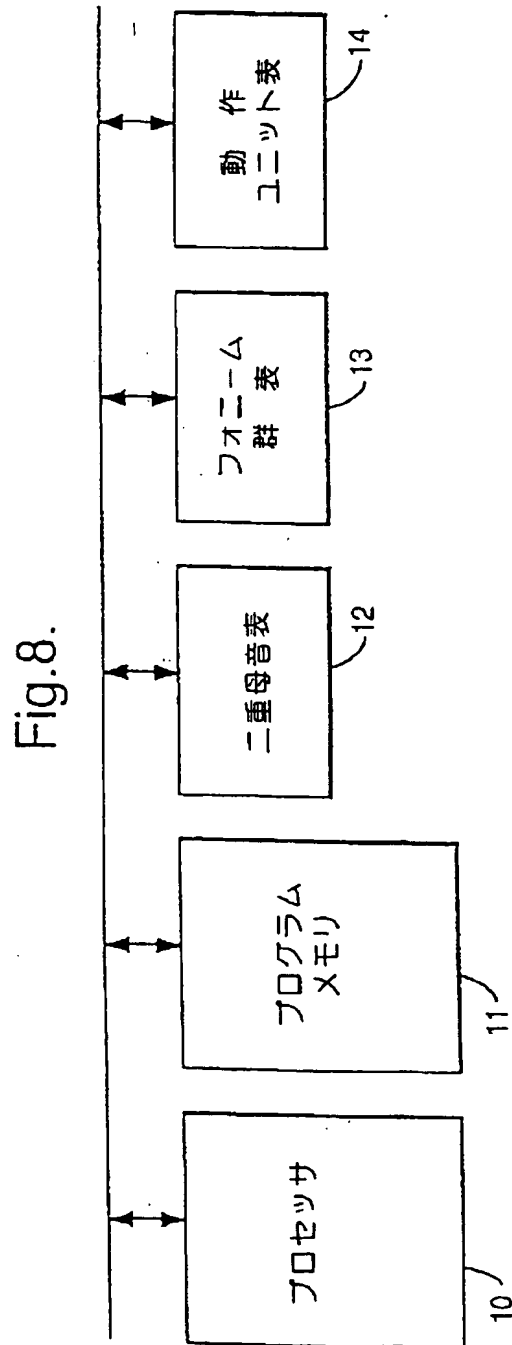


【図7】

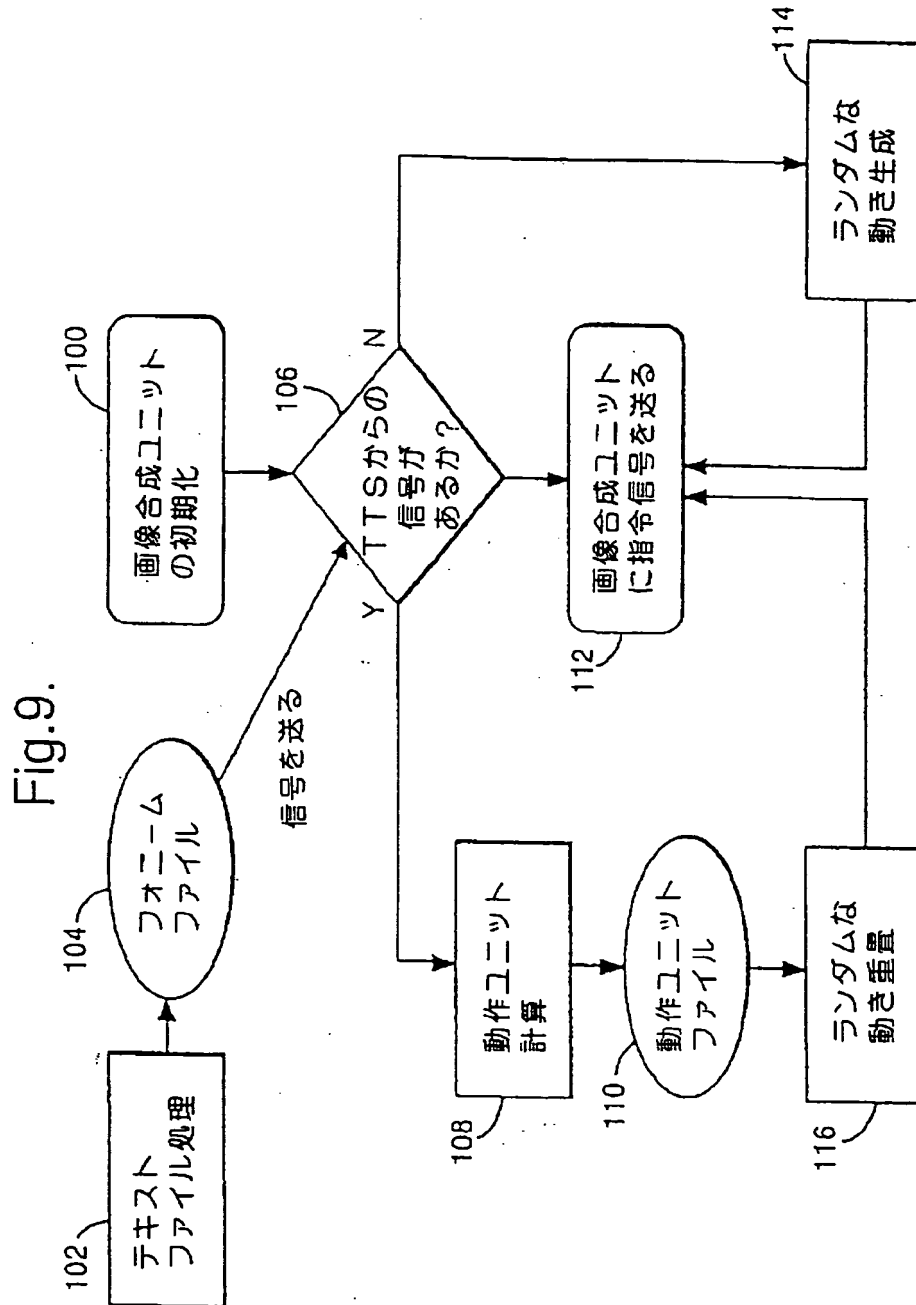
Fig.7.



【図8】



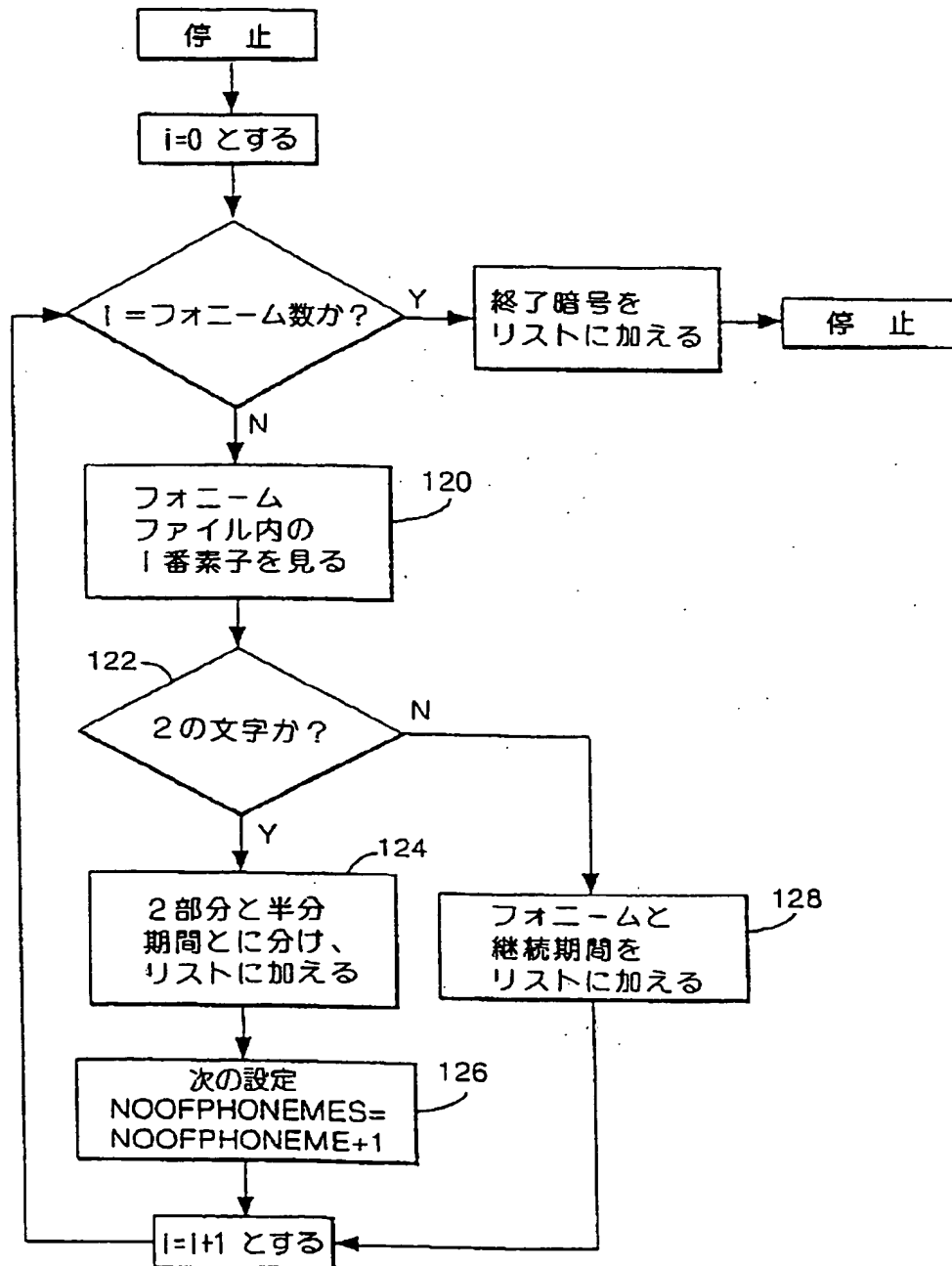
【図9】





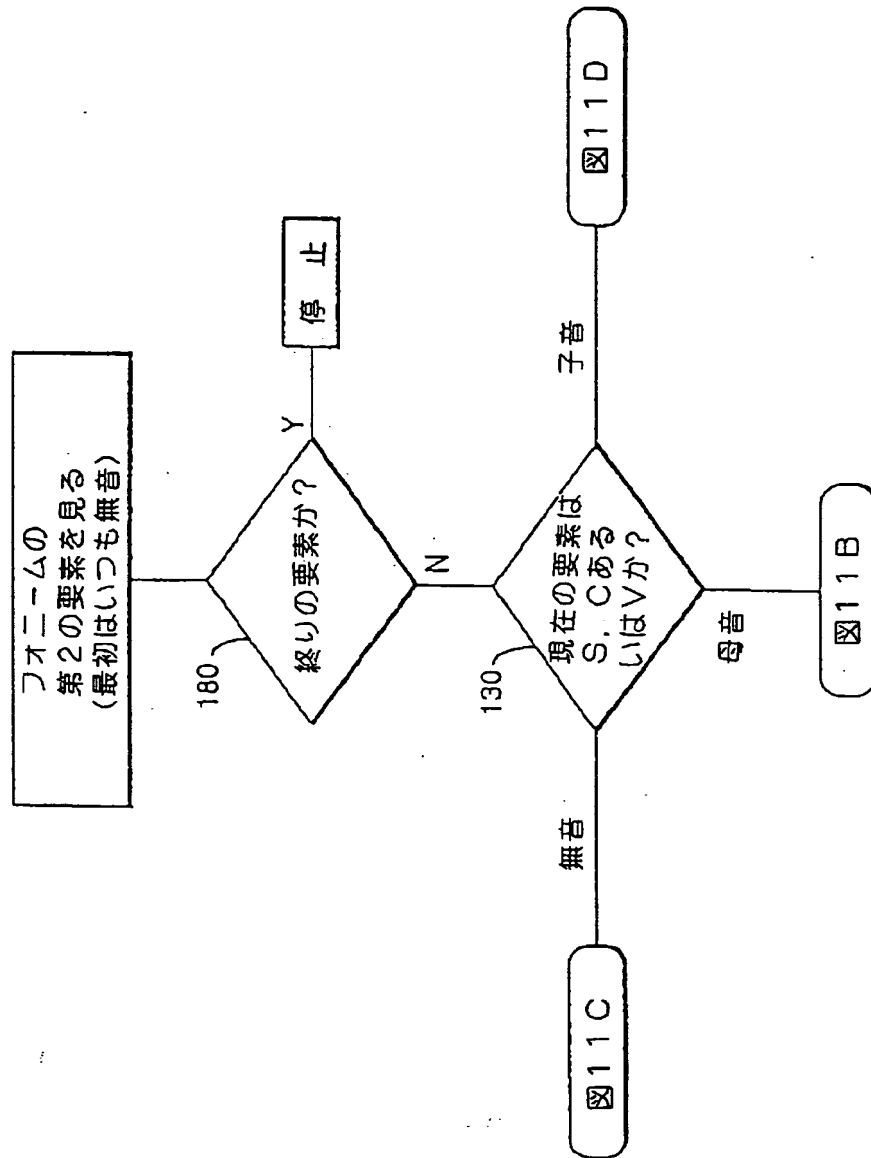
【図10】

Fig.10.

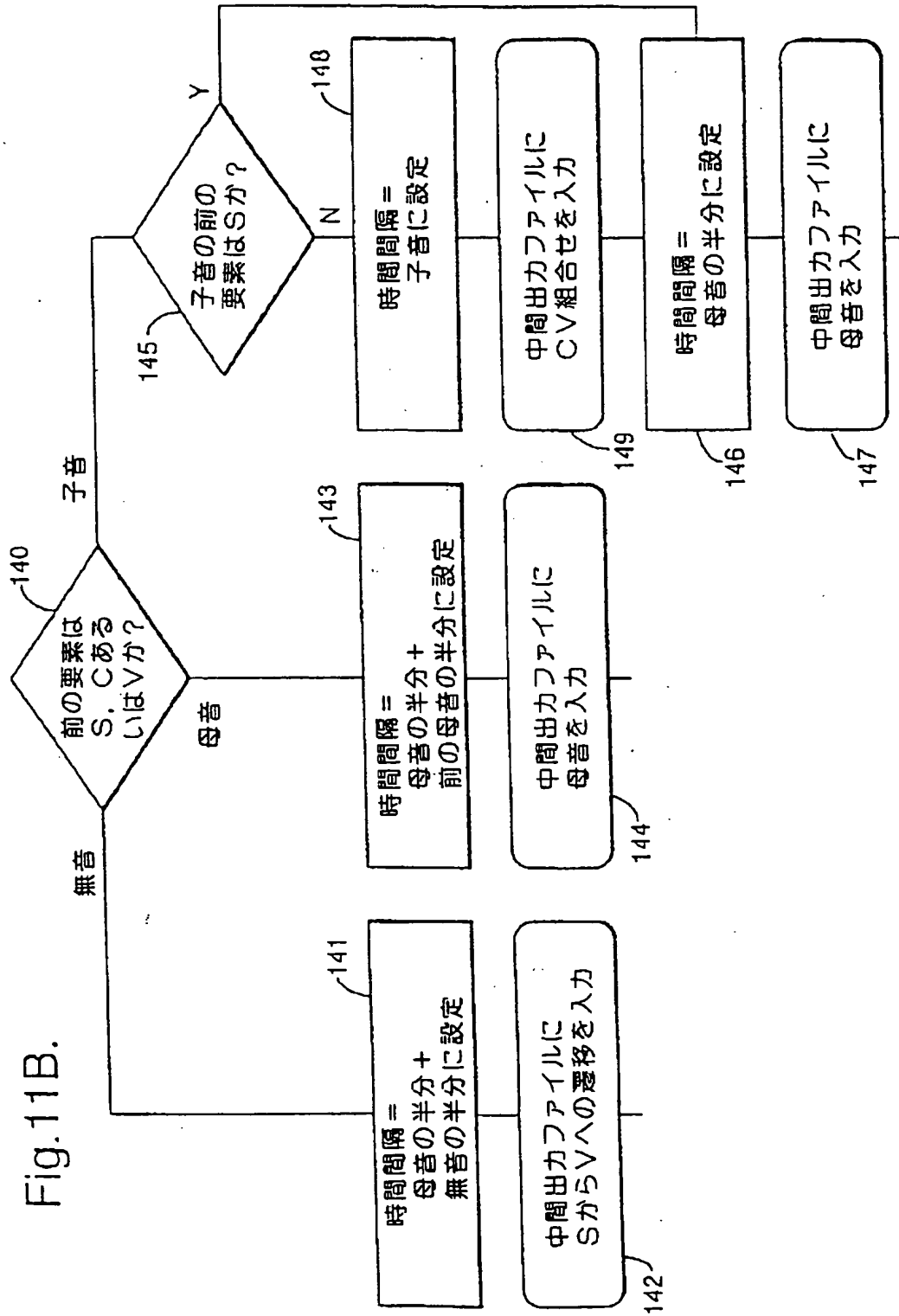


【図11】

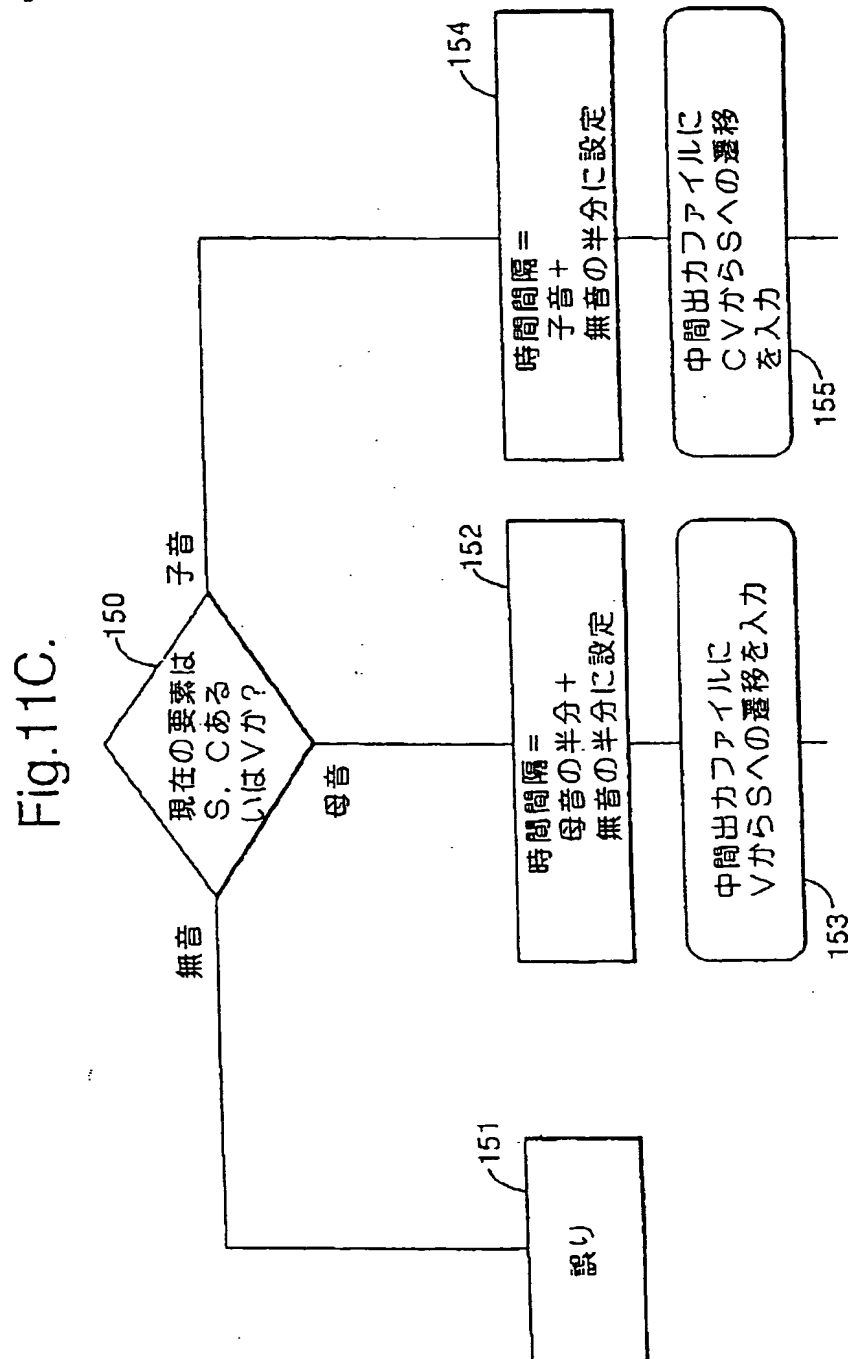
Fig.11A.



【図11】

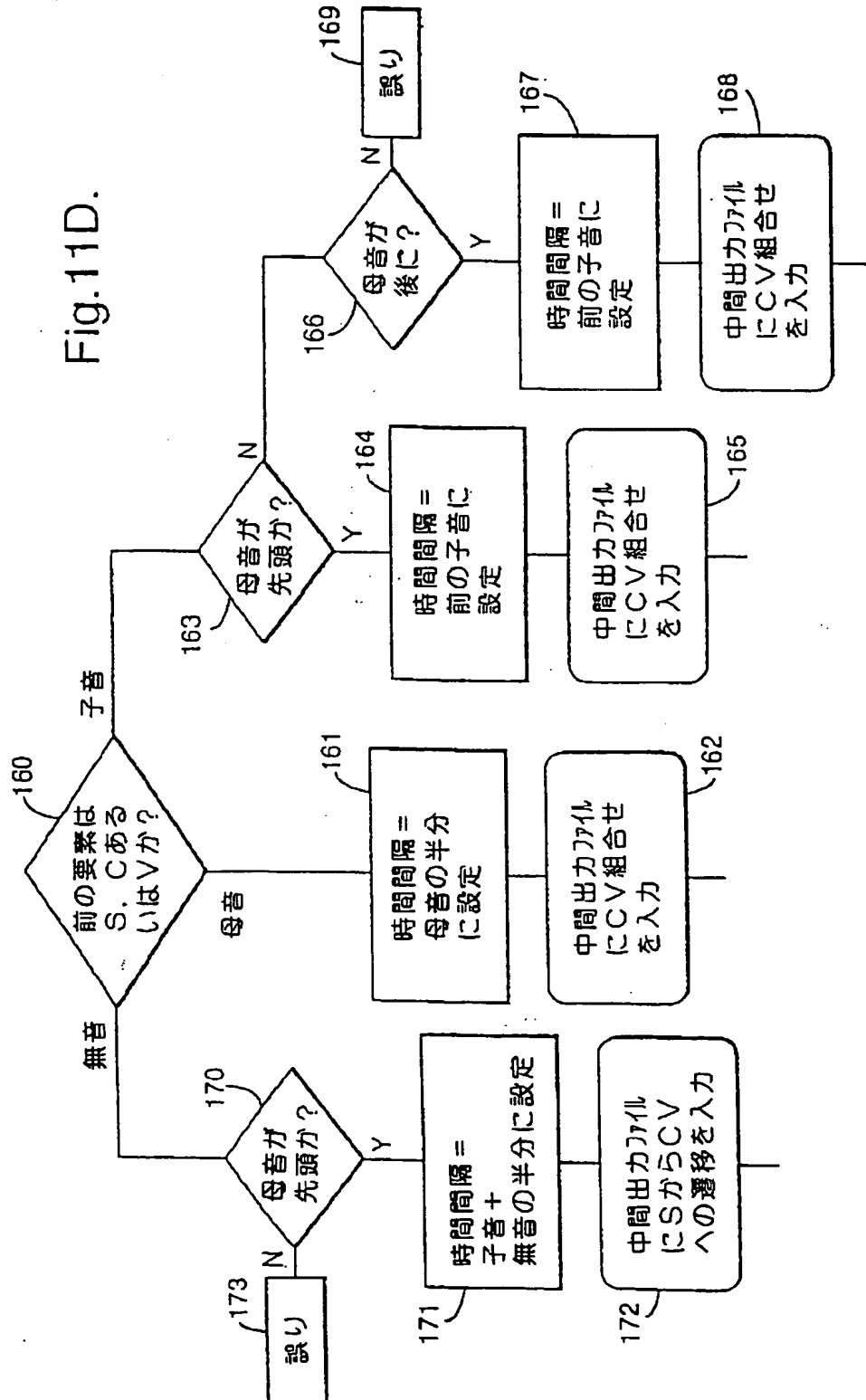


【図11】



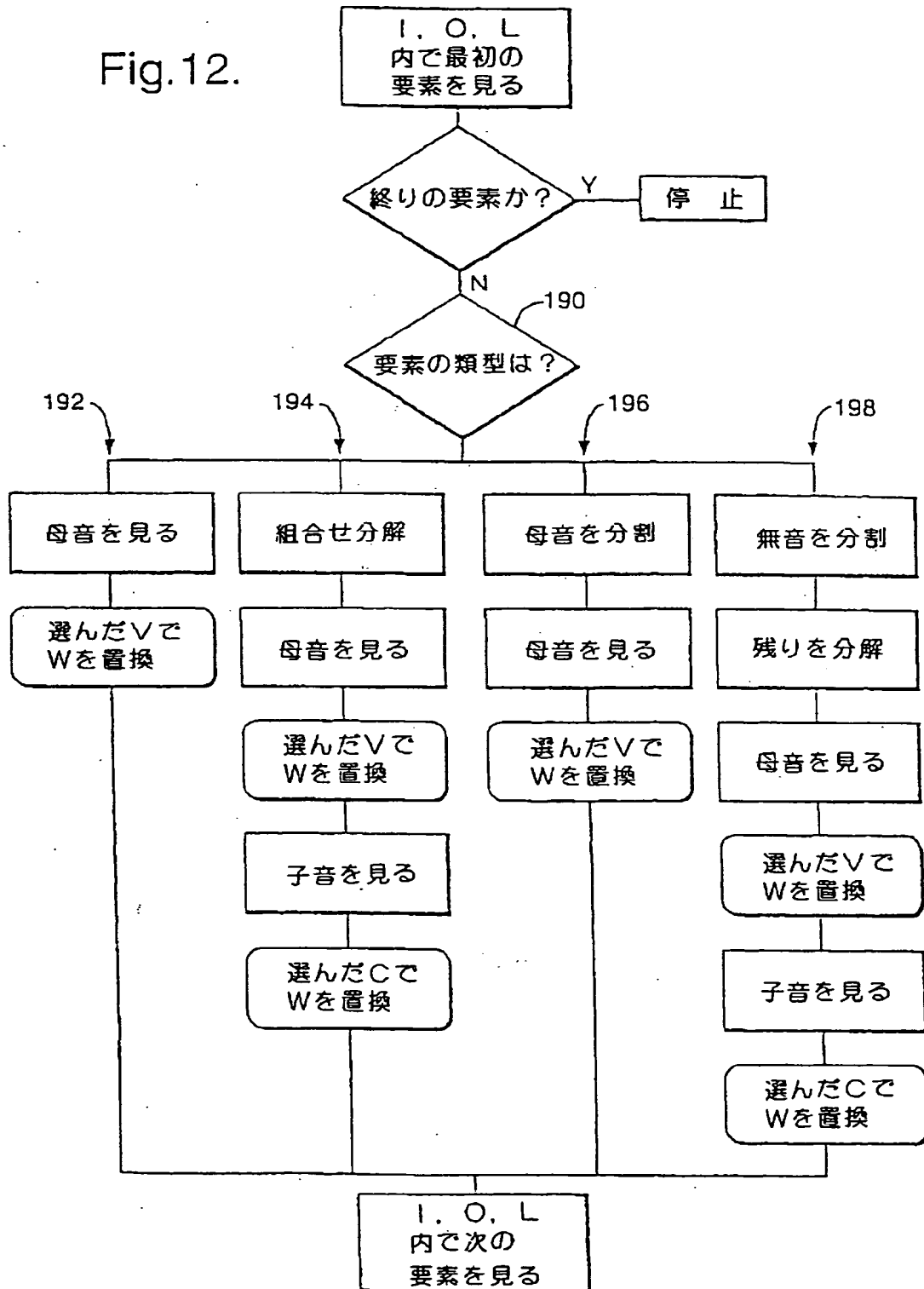
【図11】

Fig.11D.



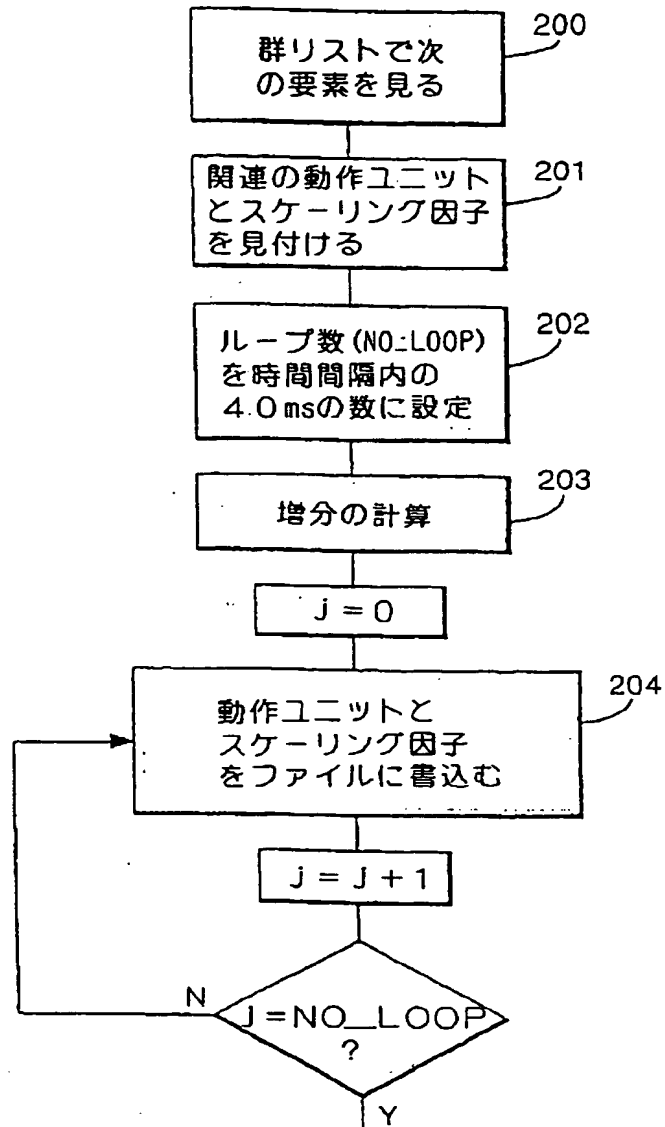
【図12】

Fig.12.



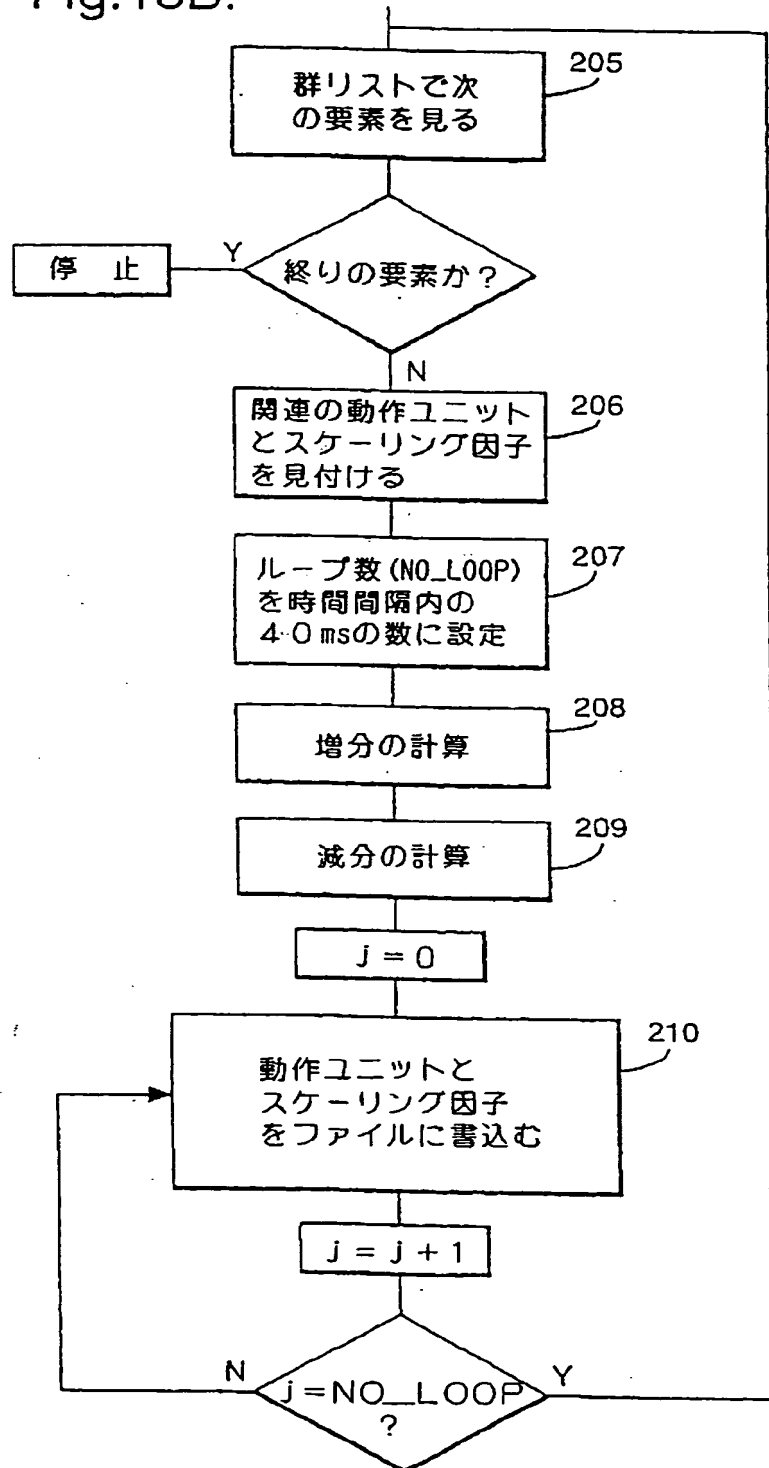
【図13】

Fig.13A.



【図13】

Fig.13B.





## 【国際調査報告】

## INTERNATIONAL SEARCH REPORT

 Intern. Application No.  
PCT/GB 97/00818

<b>A. CLASSIFICATION OF SUBJECT MATTER</b> IPC 6 G10L9/20 G06T15/78 . H04N7/26		
According to International Patent Classification (IPC) or to both national classification and IPC		
<b>B. FIELDS SEARCHED</b> Minimum documentation searched (classification system followed by classification symbols) IPC 6 G10L G06T H04N		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practical, search terms used)		
<b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b>		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	SYSTEMS & COMPUTERS IN JAPAN, vol. 22, no. 5, 1 January 1991, pages 50-59, XP000240754 SHIGEO MORISHIMA ET AL.: "A FACIAL MOTION SYNTHESIS FOR INTELLIGENT MAN-MACHINE INTERFACE" see page 50, left-hand column, paragraph 3 - page 51, right-hand column, paragraph 1 see page 52, right-hand column, paragraph 2 - page 54, right-hand column, paragraph 3; figures 1,6; tables 1-3 ---	1-3,8-11
A	US 5 313 522 A (SLAGER ROBERT P) 17 May 1994 see column 1, line 12 - column 3, line 14; claims 1,6; figures 1,2 --- -/--	1,2,9-11
<input checked="" type="checkbox"/> Further documents are listed in the continuation of box C. <input checked="" type="checkbox"/> Patent family members are listed in annex.		
* Special categories of cited documents : "A" documents defining the general state of the art which is not considered to be of particular relevance "E" earlier document but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "U" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such docu- ments, such combination being obvious to a person skilled in the art "A" document member of the same patent family		
Date of the actual completion of the international search 23 June 1997		Date of mailing of the international search report 02.07.97
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tlx. 31 651 epo nl Fax (+31-70) 340-2016		Authorized officer Greiser, N

## INTERNATIONAL SEARCH REPORT

International Application No.

PCT/GB 97/00818

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	EP 0 689 362 A (AT & T CORP) 27 December 1995 see column 1, line 14 - line 17; claims 1-4; figures 1,2 see column 1, line 35 - line 44 see column 3, line 20 - column 5, line 29 ---	1,2,9
A	GB 2 231 246 A (KOKUSAI DENSHIN DENWA CO LTD) 7 November 1990 see claims 1-5 ---	1,2,9
A	US 4 913 539 A (LEWIS JOHN P) 3 April 1990 see column 1, paragraph 1; figure 1 see column 1, line 66 - column 2, line 41; claims 9-19 -----	1,2,9

## INTERNATIONAL SEARCH REPORT

International Application No.

PCT/GB 97/00818

Patent document cited in search report	Publication date	Patent family number(s)	Publication date
US 5313522 A	17-05-94	NONE	
EP 0689362 A	27-12-95	US 5608839 A	04-03-97
		CA 2149068 A	22-12-95
		JP 8023530 A	23-01-96
GB 2231246 A	07-11-90	JP 2234285 A	17-09-90
		JP 2518683 B	24-07-96
US 4913539 A	03-04-90	NONE	

**THIS PAGE BLANK (USPTO)**